2.3 Lasso Regression / ℓ^1-regularization

Given sample points $x^{(1)}, \dots, x^{(n)} \in \mathbb{R}^p$, form the sample matrix $X = \begin{bmatrix} | & & | \\ x^{(1)} & \cdots & x^{(n)} \\ | & & | \end{bmatrix}_{p \times n}$

The Least Absolute Shrinkage and Selection Operator (lasso) refers to

$$\beta^{\text{lasso}} = \arg\min \underbrace{\|y - X^T\beta\|^2}_{f_1(\beta)} + \lambda \underbrace{\|\beta\|_1}_{f_2(\beta)}$$

where $\|\beta\|_1 = |\beta_1| + \cdots + |\beta_p|$, $\lambda > 0$.

$$f_1(\beta) = \|y - X^T\beta\|^2 = (y - X^T\beta)^T(y - X^T\beta)$$
$$= \beta^T \underbrace{X^T X}_{\text{sym. pos. semi-def}} \beta - 2y^T X^T \beta + y^T y$$

f_1 is a convex, smooth function.

$f_2(\beta) = \|\beta\|_1 = |\beta_1| + \cdots + |\beta_p|$

f_2 is a convex (HW), continuous but not differentiable function.

2.3.1 Math Prop.

Let $f : D(f) \subset \mathbb{R}^n \to \mathbb{R}$.

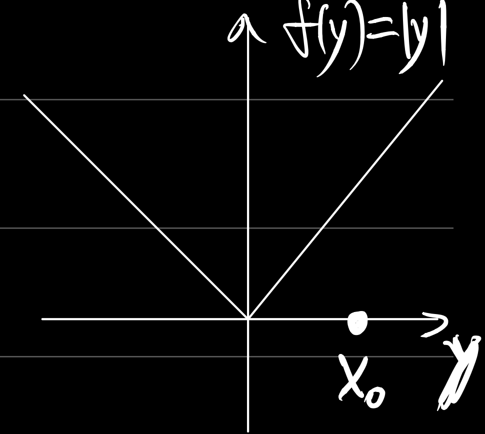
Def: A vector $s \in \mathbb{R}^n$ is called a <u>subgradient</u> of f at $x_0 \in D(f)$ if

$$f(y) \geq f(x_0) + s^T (y - x_0) \quad \text{for all } y \in D(f)$$

The set of all the subgradients of f at x_0 is called the subdifferential of f at x_0, denoted by $\partial f(x_0)$, i.e.

$$\boxed{\partial f(x_0) = \{\, s \in \mathbb{R}^n : \ s \text{ is a subgradient of } f \text{ at } x_0 \,\}}$$

Example: We compute the subdifferentiale of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$.



For $x_0 > 0$, $s \in \mathbb{R}$ is a subgradient of f at x_0,

def of subgradient

$$\Leftrightarrow \quad |y| \geq |x_0| + s(y - x_0) \quad \forall y \in \mathbb{R}$$

$$\Leftrightarrow \quad |y| \geq \underbrace{sy + (1-s)x_0}_{\text{linear in } y \text{ with slope } s \text{ and intercept } (1-s)x_0} \quad \forall y \in \mathbb{R}$$

$$\Leftrightarrow \quad \begin{cases} (1-s)x_0 \leq 0 \\ -1 \leq s \leq 1 \end{cases}$$

$$\Leftrightarrow \quad s = 1$$

Thus for $x_0 > 0$, $\partial f(x_0) = \{1\}$

For $x_0 < 0$, (exercise) $\partial f(x_0) = \{-1\}$

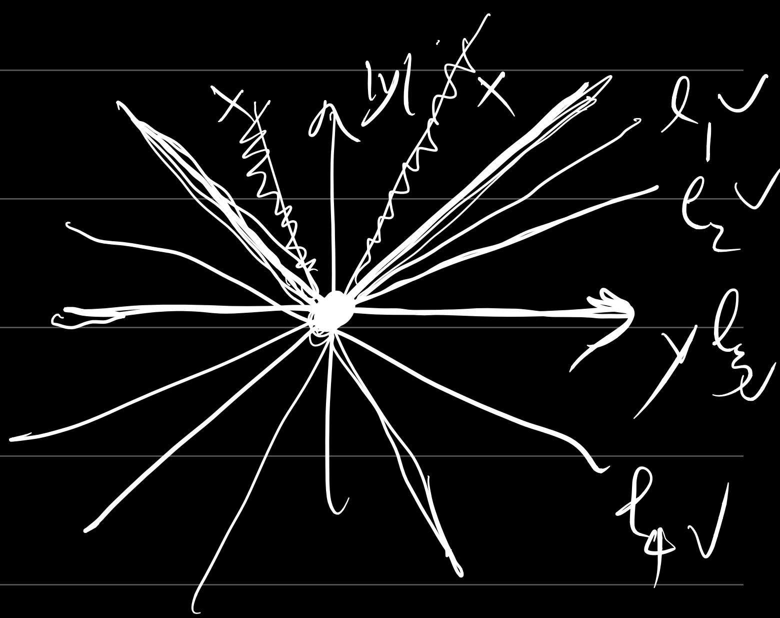
At $x_0 = 0$, $s \in \mathbb{R}$ is a subgradient of f at 0

$\overset{\text{def of subgrad}}{\Longleftrightarrow}$ $|y| \geq |\overset{=0}{x_0}| + s(y - \overset{0}{x_0}) \quad \forall y \in \mathbb{R}$

$\Leftrightarrow$ $|y| \geq \underbrace{sy}_{\text{line with slope } s \text{ and through } 0} \quad \forall y \in \mathbb{R}$

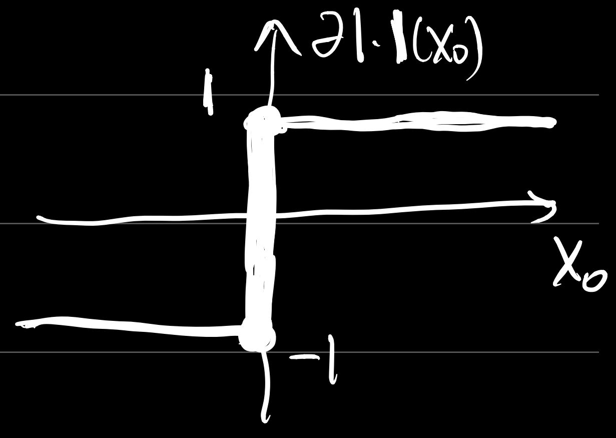
$\Leftrightarrow$ $-1 \leq s \leq 1$

$\partial f(0) = \{s : -1 \leq s \leq 1\}$



In summary:

$$\partial|\cdot|(x_0) = \begin{cases} 1 & x_0 > 0 \\ [-1, 1] & x_0 = 0 \\ -1 & x_0 < 0 \end{cases}$$



Lemma: Let $x_0 \in \mathbb{R}^n$

(1) $\partial f(x_0)$ is a closed convex set

(2) $\partial(f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0)$

where $\partial f_1(x_0) + \partial f_2(x_0) = \{ s_1 + s_2 : s_1 \in \partial f_1(x_0), s_2 \in \partial f_2(x_0) \}$

(3) If $f \in C^1$ is convex, $\partial f(x_0) = \{ \nabla f(x_0) \}$

(4) If f is convex and $\partial f(x_0) = \{ s \}$,

then f is differentiable at x_0 and

$\nabla f(x_0) = s$.